

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 09081549 A

(43) Date of publication of application: 28 . 03 . 97

(51) Int. Cl

G06F 17/21

G06F 12/00

G06F 17/28

(21) Application number: 07234287

(71) Applicant: TOSHIBA CORP

(22) Date of filing: 12 . 09 . 95

(72) Inventor: NAKAYAMA KEISUKE

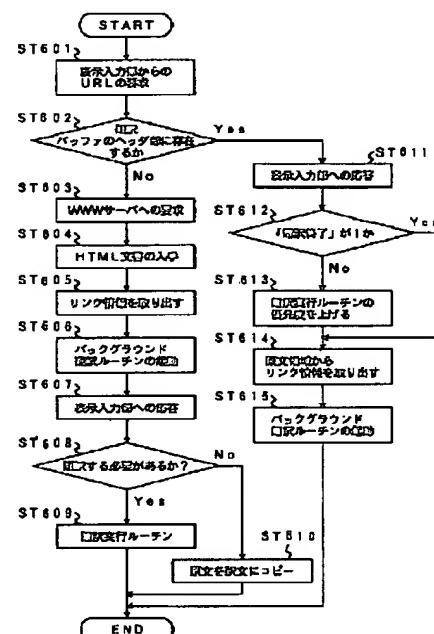
(54) TRANSLATION METHOD FOR HYPERTEXT
ACCESS

(57) Abstract:

PROBLEM TO BE SOLVED: To more shorten the response time viewed from a user in the case of access to a hypertext.

SOLUTION: The hypertext at the end of the link of hypertexts is preliminarily accessed before access request from the user, and translation of this hypertext is preliminarily started on the background. In this case, a document which may be accessed with a high possibility by the user in the hypertext in the end of the link is forecasted, and this document is preliminarily translated preferentially.

COPYRIGHT: (C)1997,JPO



Our Ref:OP1100-US

(Prior Art Reference)

Japanese Patent Laid-Open Publication No. Hei 9-81549
Laid-Open Date: March 28, 1997
Filing No. Hei 7-234287
Filing Date: September 12, 1995
Applicant: Identification Number 000003078
Kabushiki Kaisha Toshiba
Kawasaki-shi, Kanagawa-ken, Japan
Inventor : Keisuke NAKAYAMA
c/o Kabushiki Kaisha Toshiba,
Research and Development Center
Kawasaki-shi, Kanagawa-ken, Japan

Pertinent Description

(54) [Title of the Invention]

TRANSLATING METHOD FOR A HYPERTEXT ACCESS

(57) [Abstract]

[Problem]

This invention enables to shorten response time to a user when the user accesses to a hypertext.

[Means for solving the problem]

A hypertext beyond a link of a hypertext is accessed in advance before a request for an access is transmitted by a user, the hypertext beyond the link is started to be translated in a background. Further, when a translation is started in advance before the request for the access to the hypertext beyond the link of the hypertext is transmitted by the user, an estimation is made of a document having a high potentiality to be accessed by the user next in the hypertext beyond the link, and such document having the high potentiality to be accessed is translated preferentially.

[0035]

The operation of the above-described embodiment will be described hereinafter with reference to the accompanying flow chart. Fig. 6 is a flow chart showing a flow of processes in the control unit 203 of the translating unit 202. First, in step ST601, a URL request is transmitted from the display and input unit 201. In step

ST602, it is checked whether the URL request transmitted from the display and input unit 201 exists in the header section 401 of the translating buffer 204. If the URL request does not exist in the header section 401, the process advances to step ST603. If the URL request exists in the header section 401, the process advances to step ST611.

[0036]

In step ST603, the URL request received from the display and input unit 201 is transmitted to the WWW server 103. In step ST604, a document written in HTML which corresponds to the requested URL is obtained, and the URL of that document is written in a URL area of the header section 401 of the translating buffer 204. And a value of variable "translation is completed" of the header section 401 is set to 0. Further, a necessary memory area is secured in the document section 402, and the document obtained from the WWW server 103 is divided into each sentence and written in the original document area.

[0037]

In step ST605, all link information is extracted from the document written in the original document area of the document section 402 in step ST604. At this time, if a plurality of links having the same URL, i.e., links indicating the same hypertext are extracted from the original document area, these links are merged. Extracting the plurality of links leads to an estimation of a high potential access in future. In step ST606, a background translating routine (details will be described later) is activated in order to translate in advance documents beyond the all links extracted in step ST605.

[0038]

For example, if the routine shown in Fig. 6 operates on an OS which is capable of using a multitask such as an UNIX, a process such as a fork is executed for the number of times equal to the number of links, and the process is duplicated. The background translating routine is executed by the processes duplicated and newly generated. Further, for example, if the routine shown in Fig. 6 operates in a system capable of using a multithread, a new thread is generated in the number corresponding to the number of links, and the background translating routine is executed by the newly generated threads.

[0039]

At this time, the background translating routine for

processing the document beyond the link, which has been determined in step ST605 to have high potentiality to be accessed in future, is actuated with higher priority. For example, if the routine shown in Fig. 6 operates on an UNIX operating system with authorization of a super user, a renice command is executed from the parent process to the process which is newly generated by the fork and should be given higher priority.

[0040]

In step ST607, a response, that the necessary document have been obtained from the WWW server 103 and the translation will start, is made to the display and input unit 201. In step ST608, a check is made to determine in which language the document obtained in step ST604 is written, and it is checked whether the document should be translated or not. In this checking, if it is determined that the document should be translated, then the document is checked whether it is successfully consulting a dictionary, by using a plurality of languages, and it is determined that the document is written in the language with the high rate.

[0041]

If it is determined that the document should be translated, a value of variable "translation is necessary" of the header section 401 is set to 1 and the process advances to step ST609. If it is determined that the document should not be translated, the value of variable "translation is necessary" of the header section 401 is set to 0 and the process advances to step ST610.

[0042]

In step ST609, a translation executing routine (details will be described later) is executed, and a translation is divided into each sentence and written in a translation area of the document section 402 of the translating buffer 204. At this time, the link information of HTML is also written in the translation. In step ST610, the document in the original document area is copied as is in the translation area of the translating buffer 204. At this time, the value of variable "completion of translation" is set to 1.

[0043]

On the other hand, in step ST611, a response, that the necessary document had already been obtained from the WWW server 103 and the translation has already been started, is made to the display and input unit 201. In step ST612, it is checked whether the translation of

the document corresponding to the URL requested from the display and input unit 201, and which has been found in the header section 401 of the translating buffer 204 in step ST602, has already been completed or not, based on the value of variable "translation is completed" of the header section 401 of the translating buffer 204. If the value of variable "translation is completed" is 0, the process advances to step ST613. If the value of variable "translation is completed" is 1, the process advances to step ST614.

[0044]

In step ST613, regarding a document which is found in step ST612 that the translation is not completed, the process of one of translation executing routines which is translating the document is given higher priority. For example, if the routine shown in Fig. 6 operates on the UNIX operating system capable of using the multitask with authorization of the super user, a renice command is executed from the program with a process number of the process of the translation executing routine which is translating the corresponding document as an argument.

[0045]

In step ST614, all link information of the document, to which the URL request is transmitted from the display and input unit 201, is extracted from the original document area of the document section 402 of the translating buffer 204. At this time, if a plurality of links having the same URL, i.e., links indicating the same hypertext are extracted from the original document area, these links are merged. Extracting the plurality of links leads to an estimation that the potentiality to be accessed is high in future. In step ST615, the background translating routine is actuated in order to translate in advance the document beyond the all links extracted in step ST614. The method of actuating the background translating routine is the same as described for the step ST606. The process for the document which is estimated as having high potentiality to be accessed in future is also the same as described for step ST606.

特開平9-81549

(43) 公開日 平成9年(1997)3月28日

(51) Int. C1. *	識別記号	序内整理番号	F I	技術表示箇所
G 06 F	17/21		G 06 F	15/20 5 9 6 A
	12/00	5 4 7		12/00 5 4 7 H
	17/28			15/20 5 7 0 D
				15/38 Z

審査請求 未請求 請求項の数 3 O L (全 13 頁)

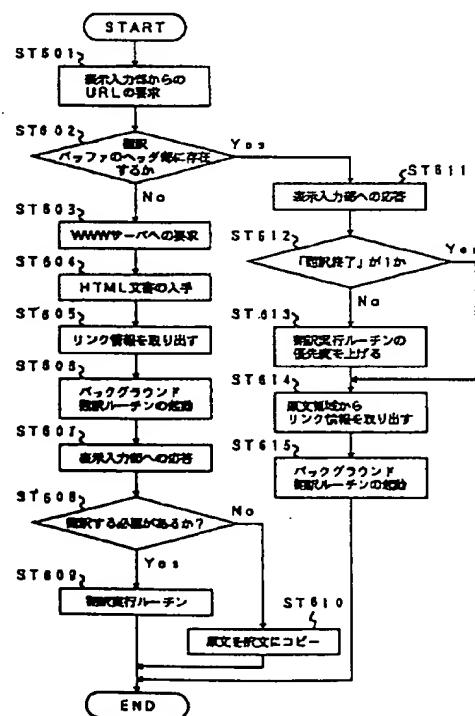
(21) 出願番号	特願平7-234287	(71) 出願人	000003078 株式会社東芝 神奈川県川崎市幸区堀川町72番地
(22) 出願日	平成7年(1995)9月12日	(72) 発明者	中山 圭介 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内
		(74) 代理人	弁理士 鈴江 武彦

(54) 【発明の名称】ハイパーテキストアクセス用翻訳方法

(57) 【要約】

【課題】ハイパーテキストにアクセスする際の、ユーザから見たレスポンス時間により短縮させることを可能にする。

【解決手段】ハイパーテキストのリンクの先にあるハイパーテキストに対してユーザからアクセス要求がある前に予めアクセスし、リンクの先のハイパーテキストをバックグラウンドで予め翻訳を開始しておく。また、ハイパーテキストのリンクの先にあるハイパーテキストに対してユーザからアクセス要求がある前に予め翻訳を開始しておく際に、リンクの先のハイパーテキストの中で次にユーザがアクセスする可能性が高い文書を予測し、アクセスする可能性が高い文書については優先して翻訳しておく。



1

【特許請求の範囲】

【請求項1】 ハイパーテキストの翻訳を行なう翻訳方法であつて、

ハイパーテキスト中のリンクが指し示すハイパーテキストの翻訳をアクセス要求がある前に予め開始しておくことを特徴とする翻訳方法。

【請求項2】 ハイパーテキストの翻訳を行なう翻訳方法であつて、

ハイパーテキスト中のリンクが指し示すハイパーテキストの中で、アクセス要求のある可能性が高いハイパーテキストを予測し、この予測したハイパーテキストの翻訳をアクセス要求がある前に予め優先的に開始しておくことを特徴とする翻訳方法。

【請求項3】 ハイパーテキストの翻訳を行ない、ハイパーテキストとその翻訳結果を保存する翻訳方法であつて、

アクセス要求があったハイパーテキストとその翻訳結果が保存されていた場合には、実際にアクセスして得られたハイパーテキストと保存されていた翻訳前のハイパーテキストとの差分のみを翻訳することを特徴とする翻訳方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、ハイパーテキストシステムにアクセスする際の言語翻訳システムに係り、特にインターネット等のネットワークを通じてWorld Wide Web (WWW) にアクセスする際のハイパーテキストアクセス用翻訳方法に関する。

【0002】

【従来の技術】 近年、インターネットの普及に伴い、World Wide Web (WWW) が注目を集めている。WWWは、インターネットを通じて、世界中の情報を自由にアクセスして情報収集を行なうことができる等の利点を持っている。

【0003】 ところが、世界中の情報に自由にアクセスできるが、諸外国の情報は外国語で書かれているものが殆どで、その際にその言語を得意としないユーザにとっては、言語の壁が、情報収集の際のボトルネックとなるという問題が発生している。

【0004】 一方、近年計算機による自然言語処理技術が高まり各種の文書処理システムが開発されている。特に、ある国の言語を他の国の言語に自動的に翻訳する機械翻訳システムなどが実用化されている。

【0005】 そこで、情報収集の際の言語の壁を取り払うために、WWW用機械翻訳システムが最近実用化された。(電子情報通信学会技術研究報告(オフィスシステム)OFS95-6 p. 31-p. 36 「快適なインターネットサーフィングを支援するWWW用機械翻訳システム」 1995年5月)

WWW用機械翻訳システムにおいては、ユーザがHyp

2

ertext Transfer Protocol (http) 等で取って来たHyper text Markup Language (HTML) 等のハイパーテキストで書かれた外国語の文書を、日本語のハイパーテキスト文書に翻訳し、ユーザに表示する。この際、元の文書のリンクの情報はそのままにして翻訳を行なう。

【0006】 一般に、HTMLのようなハイパーテキストの文書では、テキストの中に他のハイパーテキストを参照するためのタグ情報がリンクとして埋め込まれていて、そのリンクを辿ることで関連する文書にアクセスできるようになっている。

【0007】 このように翻訳を行なうことによって、インターネットを通じて入手したHTML文書等のハイパーテキスト文書を自国語で読んで情報収集を行なうことができる。また、元のHTML文書等のハイパーテキスト文書中に埋め込まれたリンクを翻訳結果のHTML文書等のハイパーテキスト文書でもそのまま保存しておく事によって、必要に応じて、原文書から同じように翻訳結果の文書からリンクの先の文書にアクセスする事ができる。

【0008】 しかしながら、通常、機械翻訳には、ある一定の処理時間が必要となる。従って、ユーザは翻訳結果が得られるまで一定時間待たなければならないという問題が生じる。

【0009】 そこで、従来のWWW用機械翻訳システムは、ユーザが翻訳処理の実行開始を指示するために翻訳ボタンを押さなくとも、原文書を表示すると同時に翻訳を始めるようになっている(事前翻訳機能)。こうすることによって、ユーザが翻訳ボタンを押す際には、文書の何割かの部分が既に翻訳されているので、ユーザに対して、翻訳ボタンが押されてから翻訳結果ができるまでの時間を短縮して翻訳結果を提示することができる。

【0010】 また、従来のWWW用機械翻訳システムでは、一度翻訳した結果を蓄積しておくことにより、再度その文書にアクセスした際には、その文書が変更されている場合のみ翻訳を行なうようになっている(蓄積選択機能)。従って、そのような際には翻訳の過程を省略することができ、ユーザの待ち時間を短縮することができる。

【0011】

【発明が解決しようとする課題】 このように従来のWWW用機械翻訳システムでは、ユーザから見たレスポンス時間を短縮するための工夫がなされている。しかしながら、ユーザから見たレスポンス時間を短縮するためにはこれだけでは不十分である。

【0012】 従来のWWW用翻訳システムでは、事前翻訳機能によって、httpで所望の文書を取って来てから文書の翻訳を始めているため、ユーザが翻訳ボタンを押す前に翻訳が始まるという利点があるものの、http

50

で文書を取って来てから翻訳ボタンを押すまでの時間が短い場合や、文書が大きい場合には、翻訳がまだ終了していないことが多い。

【0013】また、蓄積選択機能によって、過去の翻訳結果を蓄積しておくことにより、再度その文書にアクセスした際には、その文書が変更されている場合のみ翻訳を行えば良いという利点はあるものの、文書が変更されている場合には、文書全体をもう一度翻訳し直さなければならない。

【0014】本発明は前記のような事情を考慮してなされたもので、ハイパーテキストにアクセスする際の、ユーザから見たレスポンス時間をより短縮させることができた翻訳方法を提供することを目的とする。

【0015】すなわち、文書の中のリンクが指す文書についてはユーザが次にアクセスすることが予測されるので、文書の中のリンクの指す文書についてはユーザがアクセス要求を出す前に、あらかじめアクセスして翻訳を開始しておき、ユーザからみたレスポンス時間をさらに短縮することを第1の目的とする。

【0016】また、文書の中のリンクが指す文書の中でも、ユーザが次にアクセスする可能性が高い文書を予測し、ユーザがアクセス要求を出す前にあらかじめアクセスして翻訳を開始する際に、ユーザがアクセスする可能性の高い文書については優先的に翻訳しておくことによって、ユーザからみたレスポンス時間をさらに短縮することを第2の目的とする。

【0017】また、過去のアクセス履歴を参照した結果、ユーザがアクセスした文書あるいはアクセスが予測される文書が変更されていた場合でも、変更されていた文書について必要最小限の翻訳を実行することによってユーザの待ち時間を短縮することを第3の目的とする。

【0018】

【課題を解決するための手段】本発明は、第1の目的を達成するため、本発明は、WWWサーバ等のHTML等のハイパーテキストにアクセスする手段と協調し、あるいはハイパーテキストにアクセスする手段を持ち、アクセスして得られたHTML等のハイパーテキストを翻訳する手段を持つ翻訳方法であり、ハイパーテキストのリンクの先にあるハイパーテキストに対してユーザからアクセス要求がある前に予めアクセスし、リンクの先のハイパーテキストをバックグラウンドで予め翻訳を開始しておくことを特徴とする。

【0019】また第2の目的を達成するため、ハイパーテキストのリンクの先にあるハイパーテキストに対してユーザからアクセス要求がある前に予め翻訳を開始しておく際に、リンクの先のハイパーテキストの中で次にユーザがアクセスする可能性が高い文書を予測し、アクセスする可能性が高い文書については優先して翻訳しておくことを特徴とする。

【0020】またさらに第3の目的を達成するため、さ

らにアクセスしたハイパーテキストとその翻訳結果を保存する手段を持ち、ユーザからアクセス要求があったハイパーテキストとその翻訳結果が保存されていた場合には、実際にアクセスして得られたハイパーテキストと保存されていた翻訳前のハイパーテキストとの差分のみを必要最小限の翻訳作業で翻訳することを特徴とする。

【0021】このように、本発明の第1の目的を達成するための方法によれば、ハイパーテキスト中のリンクの先の文書について、その文書の翻訳をユーザがアクセス要求を出す前に始めることにより、ユーザの待ち時間を従来のWWW用機械翻訳システム等のハイパーテキストアクセス用翻訳システムよりさらに短縮することができる。

【0022】また、第2の目的を達成するための方法によれば、今後アクセスする可能性の高い文書の翻訳を予め優先度を上げて翻訳しておくことにより、その文書がアクセスされた場合に、さらにユーザの待ち時間を短縮することができる。さらに、第3の目的を達成するための方法によれば、ユーザが過去にアクセスした文書が変更されていた場合でもユーザの待ち時間を短縮することができる。

【0023】

【発明の実施の形態】以下、本発明の実施形態について図面を参照しながら説明する。図1は本発明を適用するハイパーテキストアクセス用翻訳システムの中でも代表的なWWW用翻訳システムを動作させるための構成例を示す図であり、例えばネットワークの構成図である。図1において、WWW用翻訳システム101は、ネットワークで得られた第1言語の文書を第2言語の文書に翻訳して表示するためのものであり、ソフトウェア等で実現される。なお、本実施形態では、第1言語を英語、第2言語を日本語として説明する。

【0024】翻訳結果記憶部102は、WWW用翻訳システム101が過去に翻訳した結果を格納しておくためのもので、ディスク装置等によって構成される。ディスク装置は、ネットワークファイルシステム(NFS)あるいはネットワーカードライブ等でWWW用システム101とネットワークで接続されても、WWW用翻訳システム101が動作している計算機のローカルディスクであってもよい。

【0025】WWWサーバ103は、WWW用翻訳システム101からUniform Resource Locator(URL)要求を受取り、インターネットに対してhttp要求等を出して、得られたHTML文書をWWW用翻訳システム101に返す。

【0026】図2はWWW用翻訳システム101の構成図である。図2に示すように、WWW用翻訳システム101は、表示入力部201及び翻訳部202によって構成されている。

【0027】表示入力部201は、Xウィンドウシステ

ム等のウインドウシステム上のクライアントあるいはアプリケーションとして実現され、翻訳部202で得られた翻訳結果をCRT等を用いて表示しユーザーに提示すると同時に、ユーザーからのURL要求の入力をマウスやキーボード等から受け付け、翻訳部202に送る。

【0028】翻訳部202は、制御部203と翻訳バッファ204から構成される。翻訳部202は、WWWサーバ103から送られてきたHTML等で記述された第1言語による文書を、HTMLで記述された第2言語の文書に翻訳して、その結果を表示入力部201に送ると同時に翻訳結果記憶部102に書き込む。また、表示入力部201から入力されたURL要求や、翻訳部202で生成されたURL要求を、必要に応じてWWWサーバ103に送る。

【0029】図3は表示入力部201がCRT等の画面上に表示させる内容の一例を示している。図3に示すように、表示画面中には「Document Title」「Document URL」の他、翻訳部202における翻訳結果であるハイパーテキストの一部内容

(図3では「ハイパーリンク」に関する文書)が表示されている。表示中のハイパーテキスト中には、他のハイパーテキストを指示するリンクが明示されている(図3中ではテキスト中のアンダーラインがリンクを表わす)。この、アンダーラインが付加された語句を指示することで、リンクされた他のハイパーテキストを表示させて参照することができる。

【0030】図4は翻訳バッファ204に記憶される情報を説明するための図である。翻訳バッファ204は、ヘッダ部401と文書部402が含まれる。ヘッダ部401は、複数のHTML文書についての情報を書き込むようになっており、1つ1つの文書について、そのURL、翻訳が必要であるかどうかを示す変数「翻訳必要」、その文書の翻訳が終了しているかどうかを示す変数「翻訳終了」、及びその文書が実際に書き込まれる文書部402へのポインタの4つの情報を書き込むようになっている。なお、変数「翻訳必要」は、その値が1で翻訳が必要であることを示し、0で翻訳が不要であることを示す。また、変数「翻訳終了」は、その値が1で既に翻訳が終了していることを示し、0で翻訳されていないことを示す。

【0031】文書部402は、複数のHTML文書を原文と訳文を対にして書き込めるようになっている。また文書部402は、原文と訳文が文レベルでの対応がわかる形で書き込むことができるようになっている。一般には文書部402は大きな記憶領域を必要とするので、必要に応じてダイナミックアロケーション等で計算機のメモリ上に確保するのがよい。

【0032】図5は翻訳結果記憶部102に記憶される情報を説明するための図である。翻訳結果記憶部102は、記憶部ヘッダ部501と記憶部文書部502が含ま

れる。

【0033】記憶部ヘッダ部501は、複数のHTML文書についての情報を書き込めるようになっており、1つ1つの文書について、そのURL、及びその文書が実際に書き込まれる記憶部文書部502へのポインタの情報を書き込めるようになっている。

【0034】記憶部文書部502は、複数のHTML文書を原文と訳文を対にして書き込めるようになっている。この際、原文と訳文は文レベルでの対応がわかる形で書き込むことができるようになっている。

【0035】次に、本実施形態の動作について、フローチャートを参照しながら説明する。図6は翻訳部202の制御部203における処理の流れを表すフローチャートである。まず、ステップST601では表示入力部201からURL要求が送られて来る。ステップST602では表示入力部201から送られて来たURLが翻訳バッファ204のヘッダ部401に存在するかどうかが検査される。存在しない場合はステップST603に進む。存在する場合はステップST611に進む。

20 【0036】ステップST603では表示入力部201から受け取ったURL要求をWWWサーバ103へ送る。ステップST604ではWWWサーバ103から、要求したURLの、HTML等で書かれた文書が得られると共に、翻訳バッファ204のヘッダ部401のURL領域にその文書のURLを書き込む。また、ヘッダ部401の変数「翻訳終了」の値を0にする。さらに文書部402に必要なメモリ領域を確保し、原文領域にWWWサーバ103から得られた文書を1文毎に分割して書き込む。

30 【0037】ステップST605ではステップST604で文書部402の原文領域に書き込まれた文書の中から全てのリンクの情報を取り出される。この際、原文領域から同じURLのリンク、すなわち同じハイパーテキストを指示するリンクが複数個取り出された場合にはそれらのリンクはマージされる。リンクが複数個取り出されるということは、今後アクセス可能性が高いと予測される。ステップST606では、ステップST605で取り出された全てのリンクの先の文書を予め翻訳しておくため、バックグラウンド翻訳ルーチン(詳細については後述する)が起動される。

【0038】例えば図6のルーチンがUNIX等のマルチタスクが利用可能なOSの上で動作している場合にはfork等の処理がリンクの数だけの回数実行され、処理(プロセス)が複製される。そして複製されて新たにできた処理(プロセス)でバックグラウンド翻訳ルーチンを実行する。また、例えば図6のルーチンがマルチスレッドが利用可能なシステムで動作している場合には、新しいスレッドがリンクの数だけ生成され、新たに生成されたスレッドでバックグラウンド翻訳ルーチンを実行する。

【0039】この際、ステップST605で今後アクセスされる可能性が高いと判断されたリンク先の文書を処理するバックグラウンド翻訳ルーチンは、処理の優先度を上げて起動される。例えば図6のルーチンがUNIXオペレーティングシステム上でスーパーユーザの権限で動作している場合にはfork等で新たにできた優先度を上げるべき処理（プロセス）に対し、親プロセスからreniceコマンドが実行される。

【0040】ステップST607では表示入力部201に対し、WWWサーバ103から必要な文書が得られ翻訳を開始することを応答する。ステップST608ではステップST604で入手した文書がどの言語で書かれているかが調べられ、翻訳すべきかどうかが検査される。この検査では簡単な形態素解析を行なって辞書引きに成功する単語の割合を複数の言語で調べ、その割合が高かった言語で文書が書かれていると判断する。

【0041】翻訳すべきと判断された場合は、ヘッダ部401の変数「翻訳必要」の値を1にしてステップST609に進む。翻訳すべきではないと判断された場合にはヘッダ部401の変数「翻訳必要」の値を0にしてステップST610に進む。

【0042】ステップST609では翻訳実行ルーチン（詳細については後述する）が実行され、翻訳バッファ204の文書部402の訳文領域に1文毎に訳文が書き込まれる。この際、訳文内にもHTMLのリンクの情報が書き込まれる。ステップST610では翻訳バッファ204の訳文領域に原文領域の文書がそのままコピーされる。この際、変数「翻訳終了」の値を1にする。

【0043】一方、ステップST611では表示入力部201に対し、既にWWWサーバ103から必要な文書が得られていて、既に翻訳が開始されていることを応答する。ステップST612ではステップST602で翻訳バッファ204のヘッダ部401中に見つかった、表示入力部201から要求があったURLの文書について既に翻訳が終了しているかどうかが、翻訳バッファ204のヘッダ部401の変数「翻訳終了」の値をもとに検査される。変数「翻訳終了」の値が0の場合には、ステップST613に進む。変数「翻訳終了」の値が1の場合にはステップST614に進む。

【0044】ステップST613では、ステップST612で翻訳が終了していないことがわかった文書について、その文書を翻訳している1つの翻訳実行ルーチンの処理の優先度が上げられる。例えば図6のルーチンがマルチタスクが可能なUNIXオペレーティングシステム上でスーパーユーザの権限で動作している場合は、該当する文書を翻訳している翻訳実行ルーチンのプロセスのプロセス番号を引数として、プログラム中からreniceコマンドが実行される。

【0045】ステップST614では表示入力部201からURL要求があった文書について翻訳バッファ20

4の文書部402の原文領域から全てのリンクの情報が取り出される。この際、原文領域から同じURLのリンク、すなわち同じハイパーテキストを指し示すリンクが複数個取り出された場合には、それらのリンクはマージされる。リンクが複数個取り出されるということは、今後アクセス可能性が高いと予測される。ステップST615ではステップST614で取り出された全てのリンクの先の文書を予め翻訳するため、バックグラウンド翻訳ルーチンが起動される。バックグラウンド翻訳ルーチンを起動する方法はステップST606で述べたものと同じである。ステップST615で今後アクセスされる可能性が高いと予測された文書に対する処理もステップST606でのものと同様である。

【0046】図7はバックグラウンド翻訳ルーチンの処理の流れを表すフローチャートである。ステップST701では取り出されたリンクの1つが示すURLの文書がWWWサーバ103に要求される。ステップST702ではWWWサーバ103から、要求したURLに応じた、HTML等で書かれた文書が得られると共に、翻訳バッファ204のヘッダ部401のURL領域にその文書のURLを書き込む。また、ヘッダ部401の変数「翻訳終了」の値を0にする。さらに文書部402に必要な記憶領域を確保し、原文領域にWWWサーバ103から得られた文書を1文毎に分割して書き込む。

【0047】ステップST703ではステップST702で入手した文書がどの言語で書かれているかが調べられ、翻訳すべきかどうかが検査される。この際には図6のステップST608での手法が用いられる。

【0048】翻訳すべきと判断された場合は、ヘッダ部401の変数「翻訳必要」の値を1にしてステップST704に進む。翻訳すべきではないと判断された場合にはヘッダ部401の変数「翻訳必要」の値を0にしてステップST705に進む。

【0049】ステップST704では翻訳実行ルーチンが実行され、翻訳バッファ204の文書部402の訳文領域に1文毎に訳文が書き込まれる。この際、訳文内にもHTMLのリンクの情報が書き込まれる。ステップST705では翻訳バッファ204の訳文領域に原文領域の文書がそのままコピーされる。この際、変数「翻訳終了」の値を1にする。

【0050】図8は翻訳実行ルーチンの処理の流れを表すフローチャートである。ステップST801では翻訳しようとしている文書のURLが図5の翻訳結果記憶部102の記憶部ヘッダ部501に存在するかどうかが調べられる。その文書のURLが翻訳結果記憶部102の記憶部ヘッダ部501に存在する場合には、既にその文書は過去に1回以上翻訳されている。この場合にはステップST802に進む。文書のURLが翻訳結果記憶部102の記憶部ヘッダ部501に存在しない場合にはステップST805に進む。

【0051】ステップST802では翻訳しようとしている文書、すなわち翻訳バッファ204の文書部402の原文領域にある文書と、その文書と同じURLを持つ翻訳結果記憶部102の記憶部文書部502の原文領域にある文書との間に差分があるかどうかが1文ずつ検査される。1文でも差分がある場合にはステップST803に進む。文書全体が同じであった場合にはステップST806に進む。

【0052】ステップST803では例えば特開平2-118774号に記載された方法によって、翻訳しようとしている文書、すなわち翻訳バッファ204の文書部402の原文領域にある文書と、その文書と同じURLを持つ翻訳結果記憶部102の記憶部文書部502の原文領域にある文書との差分が文単位で翻訳される。その際、1文翻訳される毎に、訳文が翻訳バッファ204の文書部402の訳文領域に書き込まれる。

【0053】差分がなく既に翻訳されている文については翻訳結果記憶部102の記憶部文書部502の訳文領域にある訳文1文が翻訳バッファ204の文書部402の訳文領域にコピーされる。文書全体の訳が得られるとステップST804に進む。

【0054】ステップST804では全文の翻訳が終了したので、その文書についての変数「翻訳終了」の値を1にすると共に、翻訳結果を翻訳バッファ204からコピーして翻訳結果記憶部102に書き込む。その際に原文も訳文と同様に翻訳結果記憶部102に書き込み、原文と訳文は記憶部文書部102に文単位の対応が取れる形で書き込む。またその文書のURLも同時に記憶部ヘッダ部501に書き込んでおく。これらの翻訳結果記憶部102への書き込みの際、文書の差分が翻訳された場合には新たな領域に書き込まれるのではなく、これまでの領域に上書きされる。

【0055】ステップST805ではステップST801で過去の翻訳結果が得られなかったので、翻訳バッファ204の文書部402の原文領域にある該当する文書の全文が翻訳される。その際、1文翻訳される毎に、訳文が翻訳バッファ204の訳文領域に書き込まれる。

【0056】ステップST806では、ステップST802で現在翻訳しようとしているURLの文書が起去に翻訳した時から変化していないことが明らかになったので、翻訳結果記憶部102の記憶部文書部502の訳文領域にあるその文書の訳文が全て翻訳バッファ204の訳文領域にそのままコピーされる。

【0057】図9はユーザからのURL入力に対する表示入力部201の処理の流れを表すフローチャートである。ステップST901ではユーザがマウスで表示入力部201の画面をクリックする、あるいはキーボードでURLを直接タイプする等の方法でURLを入力する。

【0058】ステップST902ではユーザから入力されたURLについて翻訳部202にURL要求を出す。

ステップST903では翻訳部202から、要求されたURLの文書についてWWWサーバ103から文書を入手して翻訳を始めた旨、あるいは予め翻訳が始まっていた旨、応答が帰って来る。

【0059】ステップST904では翻訳部202に要求したURLの文書について既に全文の翻訳が終了しているかが検査される。この際、翻訳バッファ204のヘッダ部401の該当するURLについての変数「翻訳終了」にアクセスし、この変数が0になつていれば翻訳がまだ途中であると判断され、ステップST905に進む。この変数「翻訳終了」が1になつていれば、翻訳が終了しているあるいは翻訳する必要がない文書だと判断されてステップST907に進む。

【0060】ステップST905では翻訳バッファ204の文書部402の訳文領域にアクセスされ、該当するURLの文書の訳文を翻訳が出来上がった文まで画面に表示される。

【0061】ステップST906では翻訳作業が進むのを待つために一定時間待機する。この際、翻訳部202が処理速度の速い計算機上にある場合は短い待機時間にし、翻訳部202が処理速度の遅い計算機上にある場合は長めの待機時間にするとよい。一定時間待機した後、ステップST904に戻る。

【0062】一方、ステップST904で全文について翻訳が終了していれば、ステップST907で翻訳バッファ204の文書部402の訳文領域にアクセスし、該当するURLの文書の訳文を全文画面に表示する。

【0063】このようにして、表示中のハイパーテキストの中のリンクが指すハイパーテキストについては、ユーザによって次にアクセスされることが予測されるので、リンクの指すハイパーテキストについてはユーザがアクセス要求を出す前に、あらかじめアクセスして翻訳を開始しておくことによって、ユーザからみたレスポンス時間をさらに短縮することができる。

【0064】また、ハイパーテキストの中のリンクが指すハイパーテキストテキストが複数ある場合に、その中でユーザが次にアクセスする可能性が高いハイパーテキストを予測し、ユーザがアクセス要求を出す前にあらかじめアクセスして翻訳を開始する際に、ユーザがアクセスする可能性の高いハイパーテキストについては優先的に翻訳しておくことによって、ユーザからみたレスポンス時間をさらに短縮することができる。

【0065】また、過去のアクセス履歴を参照した結果、ユーザがアクセスしたハイパーテキスト、あるいはアクセスが予測されるハイパーテキストが変更されていた場合でも、変更されていた差分の部分についてのみ翻訳を行なうことで、必要最小限の翻訳ですみ、ユーザの待ち時間を短縮することができる。

【0066】なお、本発明は前述した実施形態に限定されるものではない。例えば図2でWWW用翻訳システム

を翻訳部202だけで構成し、表示入力部201は別にして翻訳部202とネットワークで結ぶようにしてもよい。そのような構成にすると、家庭等のエンドユーザは表示入力部201だけを家庭のマシンに持ち、比較的大きなネットワーク上の計算機に存在する翻訳部202と電話回線やISDN回線等で結んで、システムを構成することもできる。

【0067】また、例えば図1でWWW用翻訳システム101と、翻訳結果記憶部102をユーザの近くに置いて、WWWサーバ103とは物理的に離れた場所に設置し、電話回線やISDN回線等を結んでも良い。その際、WWW用翻訳システム101とWWWサーバ103の間は、Point to Point Protocol (PPP) 等のプロトコルで接続することができる。

【0068】また、逆に図1の全ての構成要素を1つの計算機内で実現してもよい。また、アクセスする可能性の高いハイパーテキストを、そのハイパーテキストのアクセス頻度から判断してもよい。

【0069】また、本発明は上で具体的に説明したWWWアクセス用に限定されたものではなく、ハイパーテキストアクセス用に一般的に利用可能である。また、本発明は、物理的な計算機の構成 (CPUが1つ、あるいは複数等) に依存するものではないが、特に本発明の処理方法を並列計算機で実行した場合、あるいはネットワークを用いた分散処理で実行した場合等、すなわち複数のCPUを同時に使用できる場合に、その効果がより大である。

【0070】

【発明の効果】以上詳述したように本発明によれば、ハイパーテキスト中のリンクの先の文書についてはユーザがアクセス要求を出す前に、あらかじめアクセスして翻訳を開始しておくことによって、ユーザから見たレスポンス時間を従来のWWW用翻訳システム等のハイパーテキストアクセス用翻訳システムよりさらに短縮することができる。その際に、ユーザが次にアクセスする可能性が高い文書を予測し、アクセスする可能性の高い文書については翻訳処理の優先度を上げることによって、さらにレスポンス時間を短縮することができる。また、過去のアクセス履歴を参照した場合に、ユーザがアクセスし

た文書あるいはアクセスする可能性の高い文書が変更されていた場合でも、必要最小限の翻訳を行なうことによって、ユーザの待ち時間を従来のWWW用翻訳システム等のハイパーテキスト用翻訳システムよりさらに短縮することができる。また、翻訳が全部終了していない場合でも、翻訳できた部分から順に表示することによって、ユーザの待ち時間を短縮することができる。

【図面の簡単な説明】

【図1】本発明を適用するWWW用翻訳システムを動作させるためのネットワークの構成図。

【図2】図1中に示すWWW用翻訳システム101の構成図。

【図3】図1中の表示入力部201が画面上に表示させる内容の一例を示す図。

【図4】図1中の翻訳バッファ204に記憶される情報を説明するための図。

【図5】図1中の翻訳結果記憶部102に記憶される情報を説明するための図。

【図6】図1中の翻訳部202の制御部203における処理の流れを表すフローチャート。

【図7】本実施形態におけるバックグラウンド翻訳ルーチンの処理の流れを表すフローチャート。

【図8】本実施形態における翻訳実行ルーチンの処理の流れを表すフローチャート。

【図9】本実施形態におけるユーザからのURL入力に対する表示入力部201の処理の流れを表すフローチャート。

【符号の説明】

101…WWW用翻訳システム

30 102…翻訳結果記憶部

103…WWWサーバ

201…表示入力部

202…翻訳部

203…制御部

204…翻訳バッファ

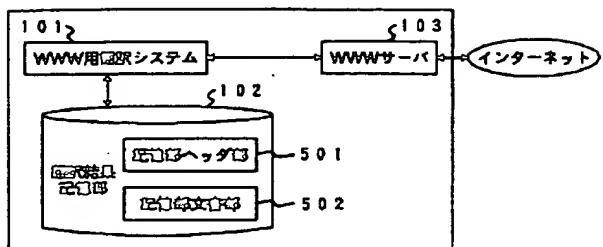
401…ヘッダ部

402…文書部

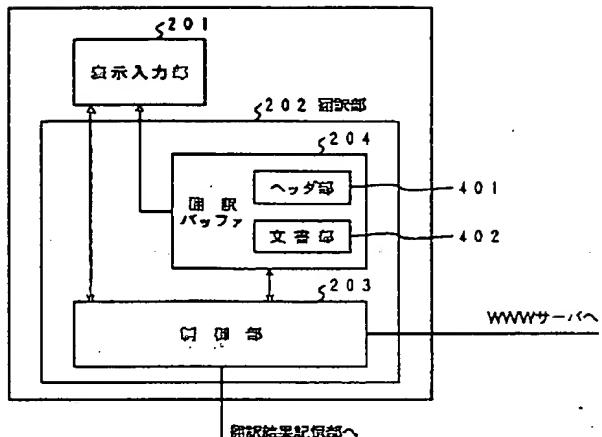
501…記憶部ヘッダ部

502…記憶部文書部

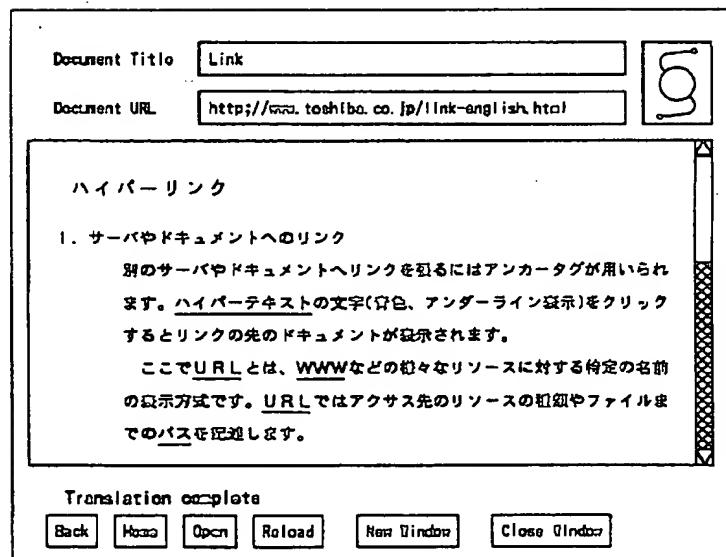
【図1】



【図2】



【図3】



【図4】

URL	OK	NG	?
http://eulca.nict.gov/Star.html	1	1	0
http://staci.edu/education.html	1	0	0
http://ccu.kyoto-u.ac.jp/index.html	0	1	0
http://...	1	1	0

2401 ヘッダ部

日本語	訳文領域
1 Jupiter is a large planet.	木星は太陽を凌ぐだ。
2 Saturn is also large.	土星もまた太りい。
...	...
18 The moon is near from the earth.	月は地球から近い。
1 Children are interested in stars.	子供達は星に興味がある。
2	...
3	...
7 They want to visit the moon.	_____
1 京大の構内	京大の構内
2	...
3	...
15 京大の近くには二寺守がなります。	京大の近くには二寺守がなります。

2402 文章部

【図5】

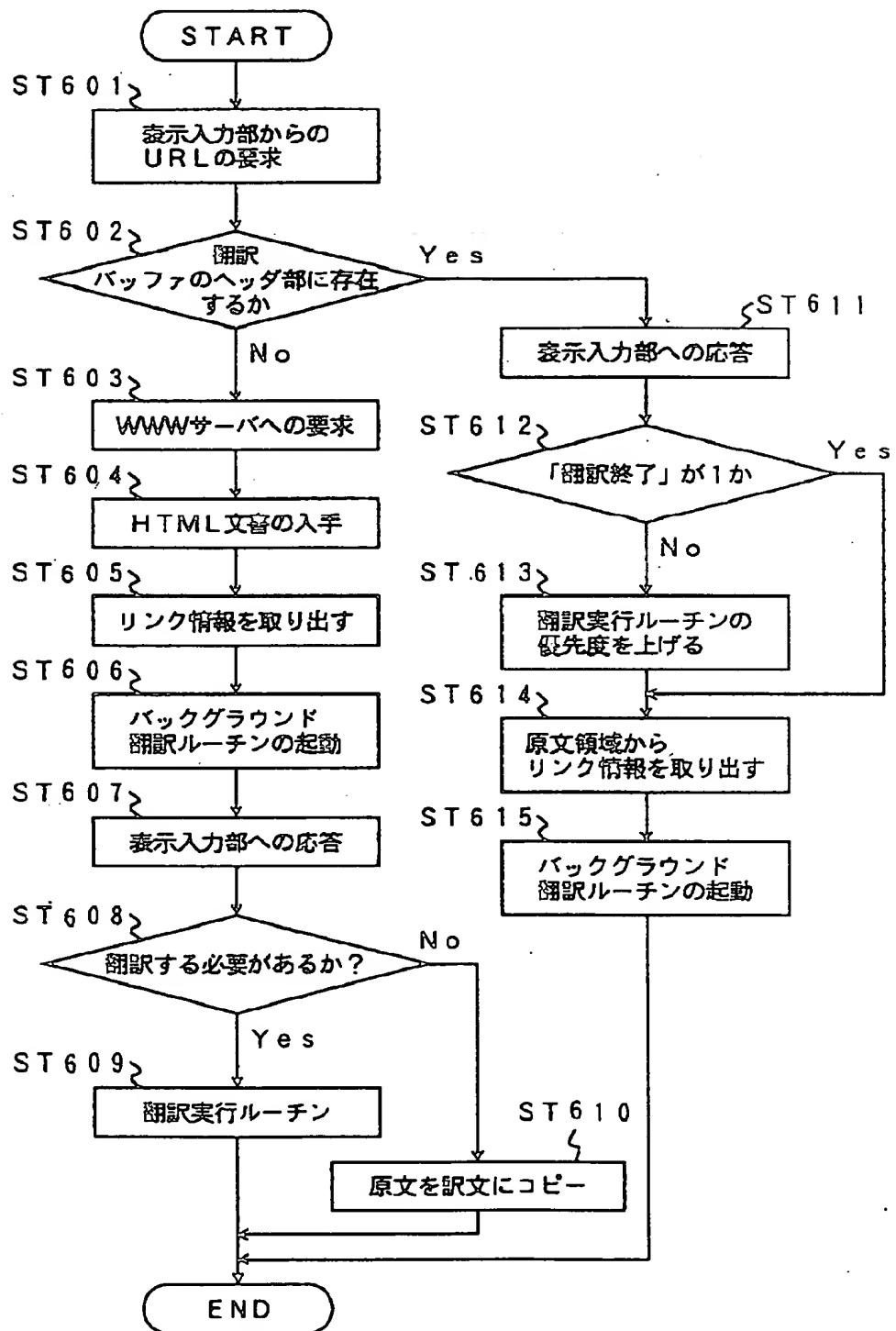
URL	?
http://cdla.ait.ca/Info.html	←
http://ccu.tohoku.ac.jp/index-eng.html	←
http://ccu.kyoto-u.ac.jp/index.html-e	←
http://...	←

501 国際化ヘッダ部

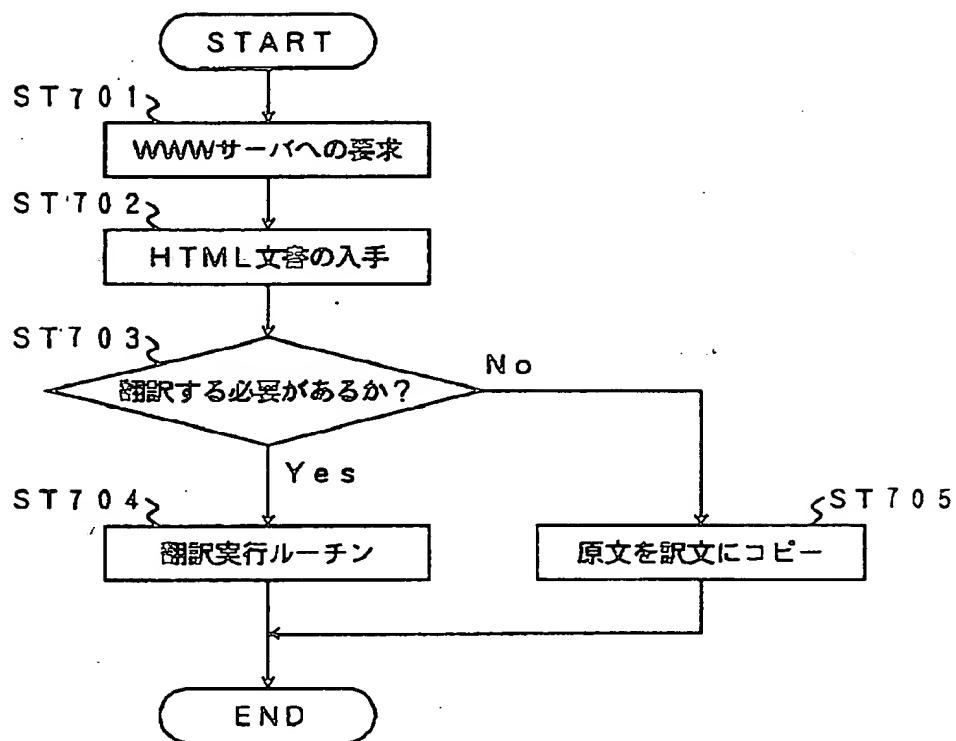
日本語	訳文領域
1 Welcome to MIT Media Lab.	MITメディアラボへようこそ。
2 MIT is located in Boston.	MITはボストンにあります。
3
4
17 I hope you will visit MIT someday.	いつかあなたMITを訪問することを希望します。
1 This is TOSHIBA's home page.	これは東芝のホームページです。
2
3
15 Thank you and good by.	ありがとうございますそしてさようなら。
1 Welcome to Kyoto Univ.	京都大学へようこそ。
2
3
8 Gionzuiji is near from Kyoto Univ.	祇園寺は京都大学の近くです。

502 国際化文書部

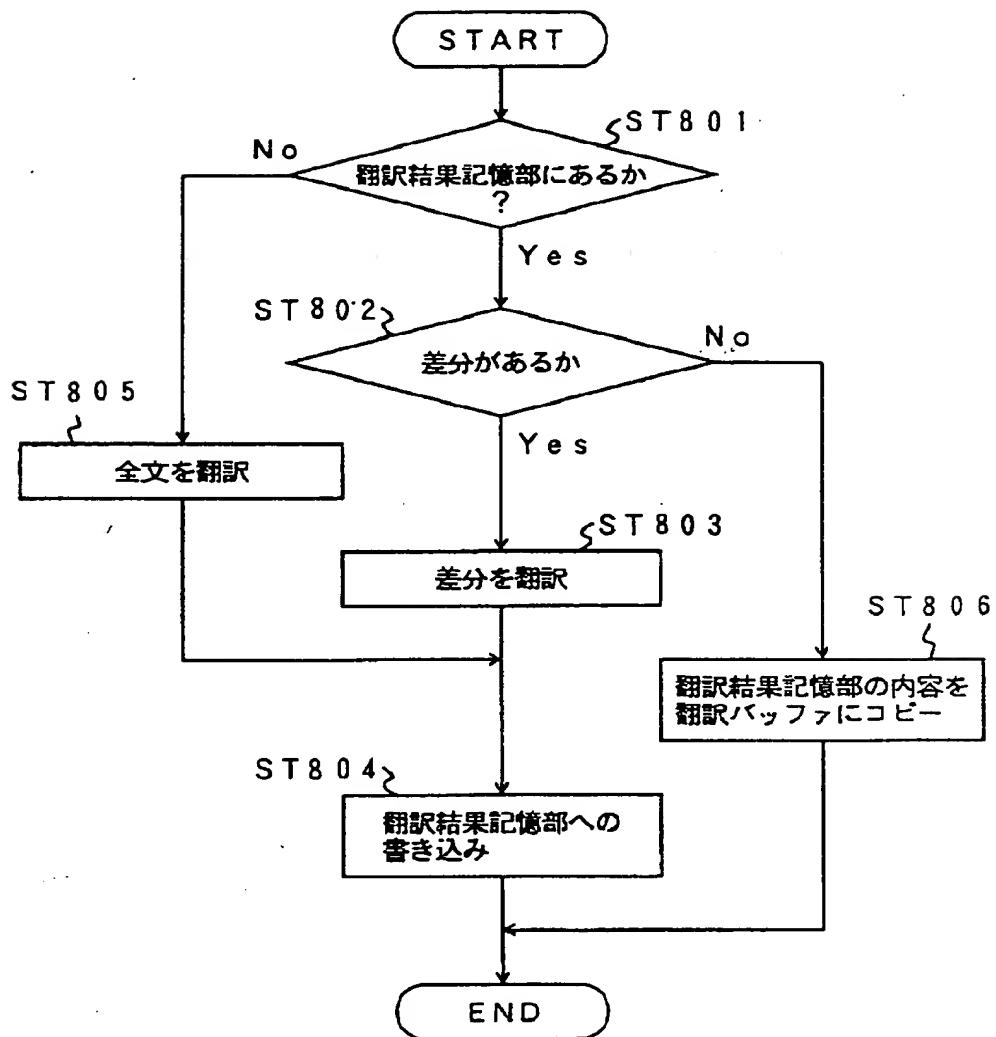
【図6】



【図7】



[图 8]



【図9】

